Article

TIG Stainless Steel Molten Pool Contour Detection and Weld Width Prediction Based on Res-Seg

Yiming Wang [†], Jing Han ^{*,†}, Jun Lu, Lianfa Bai and Zhuang Zhao ^{*}

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China; WangYiming@njust.edu.cn (Y.W.); lujunchenhao@njust.edu.cn (J.L.); blf@njust.edu.cn (L.B.)

^{*} Correspondence: eohj@njust.edu.cn (J.H.); zhaozhuang@njust.edu.cn (Z.Z.);
Tel.: +86-151-9599-5329 (Z.Z. & J.H.)

[†] Yiming Wang and Jing Han are co-first authors.

Received: 14 October 2020; Accepted: 8 November 2020; Published: 10 November 2020

Abstract: As the basic visual morphological characteristics of molten pool, contour extraction plays an important role in on-line monitoring of welding quality. The limitations of traditional edge detection algorithms make deep learning play a more important role in the task of target segmentation. In this paper, a molten pool visual sensing system in a tungsten inert gas welding (TIG) process environment is established and the corresponding molten pool image data set is made. Based on a residual network, a multi-scale feature fusion semantic segmentation network Res-Seg is designed. In order to further improve the generalization ability of the network model, this paper uses deep convolutional generative adversarial networks (DCGAN) to supplement the molten pool data set, then performs color and morphological data enhancement before network training. By comparing with other traditional edge detection algorithms and semantic segmentation network, it is verified that the scheme has high accuracy and robustness in the actual welding environment. Moreover, a back propagation (BP) neural network is used to predict the weld width, and a fitting test is carried out for the pixel width of the molten pool and its corresponding actual weld width. The average testing error is less than 0.2 mm, which meets the welding accuracy requirements.

Keywords: semantic segmentation; contour extraction; molten pool image; deep learning; weld seam width; neural network

1. Introduction

During the welding process, molten metal drops onto the base metal to form a liquid pool called the molten pool. A contour is the most basic visual morphological feature in the shape of a molten pool, and the research of welding quality control based on molten pool contour extraction [1] has made great progress. Suga et al. [2] used edge positions detected by longitudinal and horizontal scanning lines, the shape of a molten pool can be estimated. Yu et al. [3] proposed an improved edge detection algorithm based on Canny edge detector and applied it to steel plate defect detection. Li et al. [4] improved the basic model of computer vision (CV) active contour model and made it work well on a variety of images. Chen et al. [5] made improvements on the gradient operator and applied it to detect texture and edge of high temperature solidified metal. However, due to the influence of the welding process and materials, the uneven gray distribution and arc reflection on the surface of the molten pool area easily appear in molten pool images [6]. As shown in Figure 1, when the front edge of the molten pool is covered by a welding arc or the brightness saturation area on the surface of the molten pool has an impact on the rear edge of the molten pool, at this time, it is difficult to extract the accurate molten pool contour with the traditional image algorithm.

Figure 1. The effect of traditional image algorithm: (a) molten pool image; (b) contour extracted by traditional algorithm; (c) contour superimposed on molten pool image.

In recent years, with the rapid development of deep learning, it has been widely used in various industrial fields [7–9], including welding process. As one of the key problems in computer vision, semantic segmentation has aroused great interest among researchers. Semantic segmentation has made breakthroughs in many fields, and the main semantic segmentation networks include: ENet [10], SegNet [11], Fully Convolutional Networks (FCN) [12], and Unet [13]. With the support of a large data set, these networks can obtain valid results in target segmentation tasks [14]. This paper attempts to use a semantic segmentation network to solve the problem of molten pool contour extraction, but complex and diverse welding process parameters bring great difficulties to the production of a complete molten pool data set [15]. This leads to the weak generalization ability of the network model in the actual welding environment. How to make neural network learn the weak edge features in molten pool image better based on the limited data set thus becomes an urgent problem.

This paper proposes a network structure called ResSeg based on a residual network [16], which uses the superiority of a residual network to fuse the multi-scale features in the network. In addition, the data augmentation strategy based on a DCGAN network and color morphology is combined. In this paper, the network model is applied to the contour detection of TIG stainless steel molten pool image under various welding parameters. Finally, the accuracy of this method and the generalization ability of network model are verified.

2. Modeling Method

The device diagrams of a molten pool visual sensing system established in this paper are shown in Figure 2. It is mainly composed of a welding machine (TIG PI 350, Migatronik, Denmark), a robot arm (ERER-MA02010-A00-C, Yaskawa, Japan), a color charge coupled device (CCD) camera (Basler acA640-750uc, Ahrensburg, Germany), and a computer. Color CCD is used because it has the advantage of high dynamic range and can provide various high dynamic range visual information such as molten pool and arc. The camera is fixed on the robot arm of TIG welding machine at a certain angle, so in the collected image, the position of molten pool is basically fixed in a certain area of the image. Moreover, it is advantageous to suppress the influence of a welding arc light on the front end of a molten pool in

the image. In order to reduce the influence of overexposure, a neutral density filter (10%) is set in the front of a CCD, and protective glass is added to protect the camera lens.

Figure 2. Diagram of molten pool visual sensing system device.

The molten pool visual sensing system collects 1920×1200 -pixel size molten pool images. Because the proportion of the molten pool area in collected images is small, in this paper, a 400×400 -pixel region of interest (ROI) cutting is performed on collected images with a molten pool area as the center. This paper manually extracts the contour of molten pool area from images and convert it into binarized images as label in the data set. Because the traditional edge detection algorithm cannot meet the requirements of label making, this paper uses Photoshop (CC 2018, Adobe, San Jose, CA, USA) and MATLAB (R2019b, MathWorks, Natick, MA, USA) to make labels. The images after cutting and with the corresponding label are shown in Figure 3.

Figure 3. (a) Cropped molten pool images of various process parameters; (b) molten pool image and its corresponding label.

With the increasing requirements of segmentation accuracy in image segmentation task, the depth of network model is getting deeper. In some tasks, further increasing the depth of a network model is not helpful to improve the accuracy of segmentation, but leads to higher training error due to the problem of gradient disappearance. The network proposed in this paper uses a residual network as the basic structure, which can ensure that network layer is deepened as much as possible without making the network model unable to converge in the process of training, so as to obtain the optimal segmentation effect.

The existing public data sets for semantic segmentation network training, such as VOC2012 and COCO, have a very large capacity. The VOC2012 contains 21 categories of data, including tens of thousands of image data used for training alone, while COCO contains 80 categories of data, and the data used for training has also increased to the order of 100,000. In the application environment of this paper, the number of molten pool images collected by the molten pool visual acquisition system is limited, and the process of label making is quite complicated. In order to obtain the network model with higher robustness in the case of a limited data set, this paper use DCGAN to generate similar

images based on the real images in a data set to expand the original data set. The images generated by this method are one-to-one, corresponding to the original images. Although there are random differences, the overall shape and position of the molten pool area are similar. In this way, the label of a real molten pool image in the dataset is also the label of the generated image.

Before training, the images and samples in data set are augmented based on color and morphology, which further enhances the generalization ability of network model. The flow of the specific algorithm is shown in Figure 4.

Figure 4. Flow chart of system.

3. Data Set Supplement Based on Deep Convolutional Generative Adversarial Networks

Generative adversarial networks (GANs) [17], as a popular deep learning model in recent years, have shown their prominent position in the field of unsupervised learning from the very beginning. It is believed that this type of network will play an important role in the future. The training process of GANs can be regarded as the game between a generator and discriminator in network structure. The generator generates an image based on random noise and the discriminator determines whether the generated image is the original image. As the epochs of training increase, the image generated by generator is more and more similar to the original image, and the discriminator is harder to distinguish the authenticity of the generated image. Based on the original GANs, the DCGAN [18] has been improved by replacing the generator and discriminator in the original network with a convolutional neural network [19], which enables the network to extract deeper image features.

In this paper, DCGAN is used to generate similar data. The specific operation process is as follows:

- (1) Set the batch size of the network training to 4, and send the molten pool image in the data set to the network for training.
- (2) Suppose that the number of images in the original dataset is N , epoch (number of training rounds) = 500, the number of training times in each epoch is an integer rounded by $N/\text{batch size}$, test the images in the original dataset and save the network model after 100 times of batch size training.
- (3) After training, the final network model is used to test the molten pool image in the data set and the test results are saved.

Using the network model saved in step 2 of the above process to test the real molten pool images, the test results are shown in Figure 5. It can be seen that, with the increase of the number of training iterations, the generated image becomes clearer and closer to the molten pool image in the original data set.

Figure 5. Images generated during training (a) epoch = 10; (b) epoch = 120; (c) epoch = 300.

The test results on the original data set are shown in Figure 6. It can be seen that the image information of molten pool area in the generated image is still dominant. On the basis of the main image information, the random generated noise, overlapping mixture, and color change are mixed. The information in this part well simulates the unknown situation in an actual industrial welding environment, including the difference of molten pool shape characteristics caused by different welding process parameters, special workpiece materials, and abnormal welding conditions under strong light.

Figure 6. Images generated by the training model.

In the last step, based on the generated molten pool image, the corresponding real image and label are searched for in the original data set to complete the expansion of the data set, as shown in Figure 7.

Figure 7. Generated images, real images and their corresponding label.

4. Res-Seg Network Structure

The traditional convolution neural network has achieved many good results in the image segmentation task, but with the deepening of network layers, it may cause gradient problems, resulting in gradient disappearance or gradient explosion. Residual network solves this problem to a certain extent. Its main idea is to add skip connections in the network [16]. Compared with a traditional convolution neural network, residual network can learn deeper feature information of images while ensuring the convergence of network model. Based on this advantage, the accuracy of molten pool area segmentation can be fully guaranteed.

The main change Res-Seg makes is removing the full connection layer from the residual network and building a network structure similar to the fully convolutional networks (FCN). In convolution neural network, the output of deep convolution layer will lose a lot of detail information in the input image, which makes segmentation result rough, this situation is more common in the residual network. However, high-level feature contains rich and abstract image semantic information, including the location, approximate shape and category of segmentation target. Whether it is a high-level feature or low-level feature, it is very important to the final target segmentation result. In order to solve this problem, Res-Seg combines the feature information of different scales obtained in the process of downsampling with multi-scale fusion in the process of upsampling, and gets the target segmentation result with the same size as the input image through the operation of upsampling.

The network structure of Res-Seg constructed in this paper is shown in Figure 8, which is based on the improvement of ResNet-50. It can be seen from the Figure 8 that blocks in ResNet-50 are stacked in the form of {3, 4, 6, 3} and convolution operations are performed three times in each block.

Figure 8. Res-Seg network structure.

The right side of Figure 8 visually shows the change of output dimension in the process of downsampling and upsampling in Res-Seg. The related operations in the process of upsampling in Res-Seg are introduced in detail:

- (1) After downsampling stage, the feature with the size of $13 \times 13 \times 2048$ is obtained, which is equivalent to $1/32$ of the input image size. On the basis of this feature, the convolution operation with the kernel size of 1×1 is performed, and the feature $f_{1/32}$ with the size of $13 \times 13 \times 2$ is obtained;
- (2) If the feature $f_{1/32}$ is upsampled directly to the size of input image, the length and width of feature will be expanded by 32 times after one convolution operation, the segmentation result will be rough. Therefore, $f_{1/32}$ is first upsampled to feature with size of $25 \times 25 \times 2$;
- (3) From Figure 8, it can be found that the output feature size of block set with stack number of 6 is $25 \times 25 \times 1024$ in the process of downsampling. At this time, the feature $f_{1/16}$ with size of 25×25

$\times 2$ can also be obtained by using the convolution operation with the kernel size of 1×1 . In order to fuse multi-scale feature information, $f_{1/16}$ and feature obtained after upsampling on $f_{1/32}$ are added in corresponding dimensions;

- (4) Repeat the above operations for the feature obtained in step (3) and fuse them with the feature with size of $50 \times 50 \times 512$ outputted during downsampling process. Finally, carry out the upsampling operation to make the feature return to the size of input image, and obtain the feature map with the size of $400 \times 400 \times 2$.

The multi-scale feature fusion operation in the above upsampling process can be summarized as Equation (1), where $D_{1/k \rightarrow 1/h}(f_{1/k})$ represents the upsampling operation for feature $f_{1/k}$, and $\oplus$ represents the fusion operation between features.

$$f_{output} = D_{1/8 \rightarrow 1}(D_{1/16 \rightarrow 1/8}(D_{1/32 \rightarrow 1/16}(f_{1/32}) \oplus f_{1/16}) \oplus f_{1/8}) \quad (1)$$

In this way, the low-level and high-level features are fully fused, which effectively improves the accuracy of target segmentation [20]. Moreover, loss function of Res-Seg is designed as shown in Equation (2), where E stands for softmax function and i, j determines whether the pixel is located in the target area f_g or in the background area b_g , y_{ij} indicates the binary prediction value of the pixel, a represents the pixel ratio of the background, and b represents the pixel ratio of the target.

$$Loss = -a \sum_{i,j \in b_g} \log E(y_{ij} = 0) - b \sum_{i,j \in f_g} \log E(y_{ij} = 1) \quad (2)$$

5. Experiment and Analysis

5.1. Data Set Preparation and Network Training

The image data of molten pool used for training and testing are collected in two times. The position and angle of camera are different during two times of collection, resulting in different positions of molten pool area in image. The experimental environment is as follows: Ubuntu 16.04 LST 64-bit operating system, two NVIDIA GeForce GT1070 (8 GB) graphics cards, and Caffe deep learning architecture. In the training process, there are 1000 molten pool images in the training set, of which 700 are real TIG welding stainless steel molten pool images obtained by visual acquisition system. In addition, in order to improve the robustness of the network model, the training set also contains 300 molten pool images generated by DCGAN. There are 100 images in the test set, all of which are real molten pool images collected. The robustness test set consists of 50 images collected under different welding process parameters rather than a data set for training and testing.

This experiment is based on TIG welding process, protective gas is argon, gas flow is 25 L/min, welding wire brand is ER316L, base material is 304 stainless steel, camera acquisition frequency is 1000 Hz and exposure time is 20 μ s. Detailed welding process parameters are listed in Table 1.

Table 1. Welding parameters.

Groups	Current (A)	Welding Speed (cm/min)	Wire Feed Speed (cm/s)
1	100	12	0.3
2	100	16	0.3
3	140	12	0.6
4	140	28	1.0
5	180	12	0.6
6	180	12	1.0
7	180	20	0.9
8	180	28	1.2

In order to make the network model more robust, data augmentation is performed on molten pool images and corresponding labels in the data set before the data are sent to the network. The augmentation operation and the data set expansion strategy based on DCGAN occur in two different stages of molten pool contour extraction scheme, but both of them can enhance the data. In this paper, based on the expanded data set, the operation of data augmentation is carried out again, combining two different forms of data augmentation, which will significantly improve the robustness of network model.

The operation flow of data augmentation is shown in Figure 9 (the red arrow represents the direction of molten pool image in the process, the blue arrow represents the direction of label in the process and the black arrow represents both of molten pool image and label). The flow includes the rotation, scaling, cutting of molten pool image and label, and the color change of the molten pool image. In order to make the effect of data augmentation better and the data after operation more random, the intensity of the above operation is adjusted according to the size of the generated random number.

Figure 9. Flow chart of data augmentation process.

The specific data augmentation process is as follows:

- (1) Set the maximum rotation angle θ , the maximum zoom factor s , and the maximum cropping length and width values h and w ;
- (2) Generate a random floating number M in the range of 0–1, set $S = 2 \times M - 1$, on the basis of several parameters mentioned in (1), multiply S to control the intensity of shape change. Generate a random floating number N in the range of 0–5 to control the intensity of color transformation, including brightness, saturation, contrast, sharpness, Gaussian blur, etc.
- (3) Rotate and scale the data, and decide whether to crop and color transform the data according to the value of S . If S is bigger than or equal to 0, then crop the image and label, and change color of the cropped image based on the intensity of N . If S is less than 0, the remaining operations will not be performed.
- (4) After the above operation process, the molten pool image and corresponding label in the data set are sent to the Res-Seg network for training.

5.2. Analysis of Network Model Test Result

After 5000 training epochs, the accuracy of the network model on training set is 95.4%. The molten pool images in the test set are tested by using the saved network model. The contour of the segmentation result is extracted and superimposed on the original molten pool image, and the comparison of contour extraction effects is shown in Figure 10.

Figure 10. Test result of test data set: (a) Canny; (b) Chan-Vese (CV) model; (c) ENet; (d) Res-Seg; (e) Res-Seg + DCGAN.

As shown in Figure 10, compared with the traditional edge extraction algorithms (a) and (b), the contour extraction scheme (c)–(e) based on the convolution neural network can obtain a smooth and complete contour edge of the molten pool close to the real molten pool boundary. It can be seen from the comparison between (c) and (d) that the contour extracted by Res-Seg is more accurate than that extracted by ENet. This is mainly because the depth of Res-Seg network is much deeper than ENet, which also leads to Res-Seg being able to extract deeper image semantic information in the process of down sampling. After fusion with the image details extracted in the shallow layer, Res-Seg is more sensitive to the location, shape, and edge details of the molten pool in the image. It can be seen from (d) and (e) that the contour extraction accuracy is further improved after the data set expansion strategy based on DCGAN.

Furthermore, calculate the segmentation accuracy of target and background based on Equation (3). P_{ii} represents the pixel that is correctly classified, $P_{ij}(i \neq j)$ represents the pixel that is misclassified and k represents the total number of categories. The test results are shown in Table 2, which verify the effectiveness of Res-Seg and data set expansion strategy.

$$A_i = \frac{\sum_{i=0}^k P_{ii}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}}, i \in (0, k-1) \quad (3)$$

Table 2. Test set segmentation accuracy.

Category	Unet	ENet	Res-Seg (Based on ResNet-34)	Res-Seg (Based on ResNet-50)	Res-Seg (Based on ResNet-101)	Res-Seg (Based on ResNet-50) + DCGAN
Molten pool	89.51%	91.44%	91.94%	93.71%	93.80%	94.77%
Background	97.20%	99.30%	99.33%	99.28%	99.21%	99.32%

In order to verify the robustness of network model, this paper tests the robustness test set, the test results are shown in Table 3. It can be seen that the segmentation accuracy of molten pool area reaches 92%. The accuracy of the scheme combined with the data set expansion strategy is increased by about 2% on the Res-Seg only based on ResNet-50, and by about 7% compared with the Res-Seg only based on ResNet-101. Moreover, the segmentation effect of Res-Seg based on ResNet-101 is worse than that of ENet. This is because the ResNet-101 network structure is too deep and there are too many parameters, which causes the network to overfit the training set data and reduces the accuracy.

Table 3. Robustness test set segmentation accuracy.

Category	Res-Seg (Based on ResNet-101)	ENet	Res-Seg (Based on ResNet-50)	Res-Seg (Based on ResNet-50) + DCGAN
Molten pool	85.37%	89.27%	90.36%	92.14%
Background	99.25%	99.03%	98.39%	99.38%

Some results of the robustness test are shown in Figure 11. Compared with the molten pool image in Figure 10, some molten pool areas in the robustness test set are significantly smaller than those in the training data set. However, the scheme proposed in this paper can still accurately segment the molten pool area in molten pool image with different welding process parameters, which shows that the network model has strong robustness.

In this paper, ResNet-50 is selected as the basic network architecture of Res-Seg for the following reasons:

As shown in Table 2, the segmentation accuracy of Res-Seg improved based on ResNet-34 is not enough. The improved Res-Seg based on ResNet-101 has a high segmentation accuracy in the test set, but it has a poor performance in the robustness test and is not practical in the actual welding environment. In addition, the time consumption of the above three kinds of deep Res-Seg network model is tested, and the results are shown in Table 4.

Table 4. Time cost of Res-Seg network testing at different depths.

Category	Res-Seg (Based on ResNet-101)	Res-Seg (Based on ResNet-50)	Res-Seg (Based on ResNet-34)
Frame rate (fps)	6.8	8.3	17.2

In summary, considering the segmentation accuracy, model robustness, and algorithm efficiency, it is most reliable to choose ResNet-50 as the basic network architecture of Res-Seg. It has high segmentation accuracy, good model robustness, and engineering practicability.

Figure 11. Test result of robustness test data set: (a) ResSeg (based on ResNet-101); (b) ENet; (c) ResSeg (based on ResNet-50); (d) ResSeg (based on ResNet-50) + DCGAN.

5.3. Prediction of Weld Width Based on Back Propagation Neural Network

Since the weld seam width has a guiding significance in molten pool quality assessment, in order to verify the practicability of network model in engineering operation, this paper compare the molten pool width calculated from the contour test results with the actual weld seam width. In this paper, the width of the circumscribed rectangle of contour detection result is the pixel width in molten pool image.

The flow of weld width fitting verification is shown in Figure 12. In order to obtain the actual weld width, this paper uses the method of line structured light scanning to obtain the three-dimensional information of the weld seam. As shown in Figure 13, before welding, make marks on the stainless-steel plate. The line structured light is used to scan the marks of the formed weld seam, and the corresponding position of the molten pool image is obtained in the collected molten pool image. In this way, the calculated molten pool width is corresponding to the actual weld width.

The BP neural network is trained by using the neural network toolbox in MATLAB, and then the test samples are tested. In the experiment, the width of molten pool area in each image and the corresponding welding current, welding speed, and wire feeding speed are taken as the input of network. The welding current and welding speed will affect the welding heat input, and the heat input determines the shape of molten pool. The wire feeding speed affects the volume of welding wire entering molten pool per unit time, thus affecting the shape of molten pool. The influence of these three parameters on the weld pool is reflected in the pixel width of the molten pool. Therefore, these four parameters are taken as the input variables of the BP neural network. It is considered that they have the same effect on the weld seam width. That is to say, the number of neurons in the input layer of the network is 4. The actual weld width corresponding to the molten pool image is taken as the output, i.e., the number of neurons in the output layer of network is 1. The structure of the weld width prediction network based on the BP neural network is shown in Figure 14.

Figure 12. Flow chart of weld seam width verification.

Figure 13. Molten pool mark frame and corresponding weld seam mark.

Figure 14. Network structure of weld width prediction based on BP neural network.

There are 3200 sets of training data and 130 sets of test data that are input as the BP neural network. Figure 15 shows the error convergence in the training process, from which we know that the BP neural network reaches the convergence state in 1050 iterations of training.

The BP neural network is used to test the test data, and the curve fitting method was used as the comparison experiment. The pixel width data of molten pool area is fitted with the corresponding weld width data, and the fitting equation is used to predict the test data. The comparison between the prediction method proposed in this paper and the other three prediction methods based on curve fitting is shown in Figure 16.

Figure 15. Error convergence curve in the training process of BP neural network.

Figure 16. Comparison between molten pool pixel width and weld width of several prediction methods: (a) ordinary primary fitting; (b) Gaussian primary fitting; (c) Gaussian secondary fitting; (d) prediction method based on BP neural network.

The data curve in Figure 16 shows a stepped distribution, because the test data include molten pool images under various welding process parameters, and the details are shown in Table 5. It can be seen that it is not robust to map the weld width using only the pixel width of the molten pool, while the error of the BP neural network method is small. The results show that the input of welding

current, welding speed, wire feeding speed and pixel width of molten pool is more decisive for the results, and neural network can learn the deeper relationship between data better than the curve fitting method.

Table 5. Prediction errors of four methods under different welding process parameters.

Groups	Current (A)	Welding Speed (cm/min)	Wire Feeding Speed (cm/s)	Ordinary Primary Fitting (mm)	Gaussian Primary Fitting (mm)	Gaussian Secondary Fitting (mm)	BP Neural Network (mm)
1	120	10	1.0	0.1802	0.2817	0.3432	0.1349
2	130	10	1.2	0.1335	0.2535	0.3209	0.0972
3	130	15	1.0	0.8167	0.8392	0.7945	0.2273
4	130	15	1.2	0.4554	0.4770	0.5158	0.1474
5	140	10	1.2	1.1136	0.8795	0.6501	0.1516
6	140	15	1.0	1.0628	0.8469	0.8472	0.1825
7	140	15	1.2	0.6706	0.3961	0.4117	0.1227
8	150	10	1.0	0.7654	0.5530	0.5779	0.1806
9	150	15	1.0	0.4974	0.4034	0.3275	0.1355
10	160	15	1.2	0.7522	1.0714	1.1366	0.2165
Average error (mm)				0.6738	0.6123	0.5994	0.1864

The predicted error and average error calculated based on the test data under different groups of welding process parameters are shown in Table 5.

It can be seen from Table 5 that the accuracy of using BP neural network to predict the weld width is greatly improved compared with the traditional fitting method. The average test error of segmented data is less than 0.23 mm, and the average test error of the whole test data is less than 0.2 mm, which meets the requirements of weld width prediction accuracy. It is proven that the generalization ability of the network scheme training model proposed in this paper is reliable and has practical value in engineering.

6. Conclusions

The image of the molten pool in the TIG stainless steel welding process is collected by using the vision acquisition system developed in this paper. A semantic segmentation network Res-Seg based on the ResNet-50 network is proposed to extract the contour of the molten pool in TIG stainless steel welding. The network incorporates multi-scale deep image features, uses DCGAN to supplement the original data set, and enhances the robustness through data augmentation. The model obtained by Res-Seg proposed in this paper has high accuracy in the contour detection of a single-frame molten pool. It is a good solution to solve the problem that the weak edge of molten pool cannot be accurately detected due to arc interference or molten pool reflection.

In addition, by using a BP neural network to predict the weld width, four parameters of molten pool pixel width, welding current intensity, welding speed, and wire feeding speed are taken as input, and the actual weld width is taken as output. The average test error is less than 0.2 mm, which meets the requirements of welding accuracy. It is proved that the network model proposed in this paper has a strong generalization ability in the image segmentation of the molten pool, and can be used for the shape quality analysis in the actual welding process.

Author Contributions: Conceptualization, methodology, writing—original draft, Y.W.; investigation, writing—review and editing, Z.Z. and J.H.; supervision, L.B. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (61727802 and 61971227), the Fundamental Research Funds for the Central Universities (30920031101), and the Jiangsu Provincial Key Research and Development Program (BE2018126).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, J.; Wang, K.; Wu, S.T.; Sun, K. Two-directional synchronous visual sensing and image processing of weld pool in aluminum alloy twin arc pulsed mig welding. *J. Mech. Eng.* **2014**, *50*, 44–50. [\[CrossRef\]](#)
2. Suga, Y.; Shimamura, T.; Usui, S.; Aoki, K. Measurement of molten pool shape and penetration control applying neural network in TIG welding of thin steel plates. *ISIJ Int.* **1999**, *39*, 1075–1080. [\[CrossRef\]](#)
3. Yu, H.; Mu, P. Application of adaptive canny algorithm in edge detection of steel plate defects. *Softw. Guide* **2018**, *17*, 175–177.
4. Li, C.; Liu, Z. An improved image segmentation model based on cv model. *J. Minzu Univ. China* **2014**, *23*, 83–87.
5. Chen, Z.; Li, Y.; Chen, X.; Yang, C.; Gui, W. Edge and texture detection of metal image under high temperature and dynamic solidification condition. *J. Cent. South Univ.* **2018**, *25*, 1501–1512. [\[CrossRef\]](#)
6. Lei, K.; Qin, X.; Liu, H.; Ni, M. Weld pool edge extraction in wide-band laser cladding based on local region active contour model. *J. Optoelectron. Laser* **2018**, *29*, 516–522.
7. Bai, Y.; Lou, Y.; Gao, F.; Wang, S.; Wu, Y.; Duan, L. Groupsensitive triplet embedding for vehicle reidentification. *IEEE Trans. Multimed.* **2018**, *20*, 2385–2399. [\[CrossRef\]](#)
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* **2016**, *93*, 490–502. [\[CrossRef\]](#)
10. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv* **2016**, arXiv:1606.02147.
11. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Assist. Interv.* **2015**, 234–241.
14. Yu, W.; Long, H. A building segmentation method based on deep convolution networks for remote sensing imagery. *Comput. Technol. Dev.* **2019**, *29*, 57–61.
15. Sun, Y.; Jiang, Z.; Dong, W.; Zhang, L.; Rao, Y.; Li, S. Image recognition of tea plant disease based on convolutional neural network and small samples. *Jiangsu J. Agric. Sci.* **2019**, *35*, 48–55.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, 770–778.
17. Goodfellow, I.; Pougetabadi, J.; Mirza, M.; Xu, B.; Wardefarley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
18. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
19. Yuan, J. Learning building extraction in aerial scenes with convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2793–2798. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Huang, Y.; Liu, Y.; Ren, H. Segmentation of forest image based on fully convolutional neural network. *Comput. Eng. Appl.* **2019**, *55*, 219–224.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).